

Scene-based Stimulation in the Context of a Visual ERP BCI

Amar Enkhbat¹, Michael Tangermann¹, Jordy Thielen¹,

¹Donders Centre for Cognition, Radboud University Nijmegen, Netherlands

E-mail: amar.enkhbat@donders.ru.nl

ABSTRACT: Visual brain-computer interfaces (BCIs) provide a promising communication channel for individuals with severe motor impairments. One well-known example is the event-related potential (ERP) matrix speller, which enables users to select characters on a computer screen. However, such systems typically require users to repeatedly shift attention between a screen-based interface and the physical environment. In this work, we propose an in-the-scene BCI paradigm, in which users interact directly with real-world objects rather than symbolic representations on a screen. We evaluated the proposed paradigm in offline experiments in which healthy participants ($N = 22$) selected target objects under two conditions: a screen-based BCI and an in-the-scene laser-based BCI. Comparisons of the elicited ERPs showed differences between the two stimulation conditions, including distinct scalp activation patterns and significantly larger N1 amplitudes in the in-the-scene condition. These findings indicate that neural responses elicited by real-world object stimulation differ from those obtained with conventional screen-based BCIs. This suggests that decoding methods may require adaptation for in-the-scene BCI systems.

INTRODUCTION

A brain-computer interface (BCI) is a type of assistive technology that decodes brain activity directly, bypassing the brain's neural pathways that control other forms of muscular activity. For people with severe speech or muscular disability such as paralysis or locked in syndrome, BCI is a promising technology that could allow them to communicate with the external world, improve independence and quality of life.

One such assistive device is the classic event-related potential (ERP) matrix speller originally described by Farwell and Donchin [1]. The 26 letters of the alphabet, together with several other symbols and commands are displayed on a computer screen that serves as a keyboard. The user focuses attention to a symbol they wish to select while rows and columns of the keyboard are highlighted. Upon highlighting letters, ERPs are elicited and captured by the electroencephalogram (EEG). By analyzing these responses across multiple stimulus presentations, the BCI determines the character attended by the user.

Although the ERP speller enables users to communicate their intentions through typing, its long-term adoption has been limited [2]. Our hypothesis is that adoption could

be limited by the sometimes cumbersome screen-based selection procedure. For example, when interacting with devices or objects in a physical environment, users must repeatedly shift attention between the screen interface and the physical environment. Additionally, during communication, users are restricted to characters or options that are pre-specified on the screen interface. The speed of interacting with a real-world object (e.g., a glass of water) often requires the user to spell out the name of the object and the type of interaction desired. Because selecting a single character typically takes several seconds, describing objects such as a glass of water or actions such as drinking from it, may take over a minute. This tedious process could lead to frustration, mental fatigue and reduced long-term usability [3].

To address these limitations, we explore the concept of an *in-the-scene BCI* originally proposed by Kolkhorst et al. [4, 5]. In their study, objects in the real-world were selected directly, circumventing the need of a screen. This was achieved by placing objects around the participant and highlighted them using a single laser pointer mounted on a moving robotic arm. The authors hypothesized, that such a system might reduce cognitive workload because users do not need to repeatedly switch attention between a screen interface and the surrounding environment. Instead, object selection could occur directly by attending to the desired item. For simplicity, we refer to the in-the-scene condition as the *laser* condition for the remainder of this paper.

However, it remains unclear whether highlighting real-world objects with varying visual and reflective properties results in different ERP responses compared to conventional screen-based stimuli that are typically presented on a homogeneous background. In this study, we use data from an experiment where participants selected objects using both, a screen condition and a laser condition. Within-subjects, we compare respective evoked ERP components regarding amplitude, latency and scalp activation characteristics.

MATERIALS AND METHODS

Participants: Twenty-five healthy participants aged between 17 and 71 years (9 male, 16 female; mean age: 38.75, SD: 16.3) participated for a single session after pre-screening and providing written informed consent. Exclusion criteria included inability to understand instructions in English, a history of color blindness, neu-

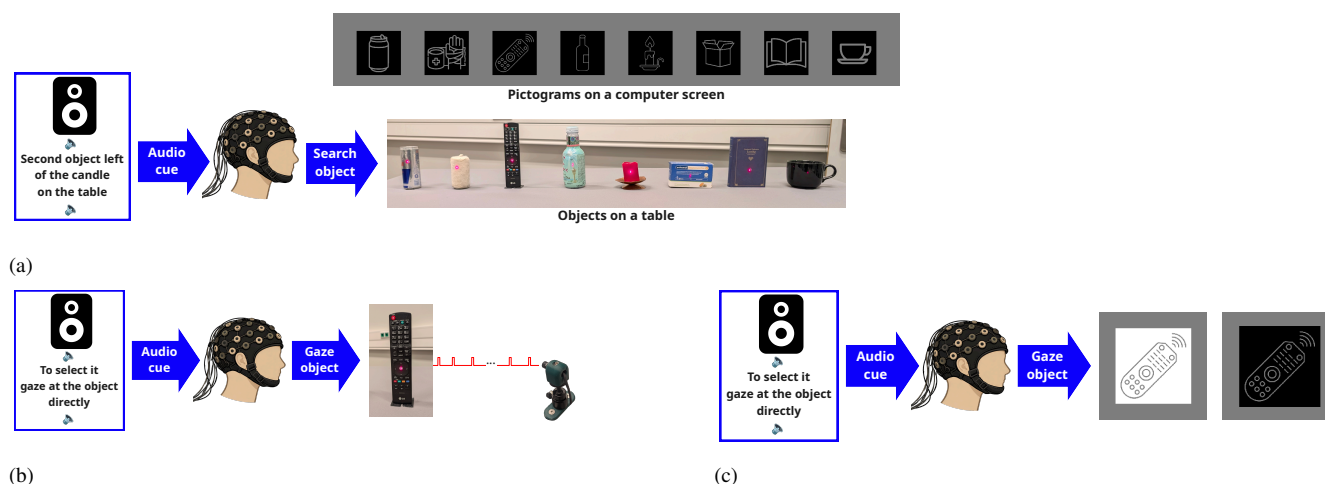


Figure 1: **Experiment overview.** (a) An audio cue instructs participants to find the target object on the table. (b) In the laser condition, participants gaze at the red laser-dot projected on the target. Highlighting is done by switching lasers on and off. (c) In the screen condition, participants gaze at the target object's pictogram on a screen. Pictograms are highlighted by alternating between white and black squares.

rological treatment, brain surgery, or epilepsy. All participants had normal or corrected-to-normal vision. Participants were compensated for their contribution to the study. The experiment was approved by the Ethical Committee of the Faculty of Social Sciences at Radboud University Nijmegen.

Materials: EEG data was recorded with 64 channel Ag/AgCl active electrodes positioned according to the 10-10 system, and amplified with a BrainProducts BrainAmp amplifier at a sampling rate of 1000 Hz. The reference electrode was placed on the nose and the ground electrode was positioned at AFz. The impedance of all electrodes was kept under 10 k Ω . Event markers were recorded and synchronized using Lab Streaming Layer (LSL). EEG acquisition and marker recording were performed using the LSL Connector for the BrainAmp amplifier.

We designed an experiment that includes the two conditions: a standard screen-based BCI and a laser-based BCI. Eight everyday objects were placed in front of the participant (can, bandage, remote, bottle, candle, box, book, cup; see Figure 1a). The objects were positioned in a line with 18 cm gap between their respective highlighting locations. Participants were seated approximately 1.5 m from the objects.

Eight laser pointers were placed to the left of the participant. Each laser was manually aligned to point at the center of mass of each object. The lasers were transistor-to-transistor logic (TTL) laser diodes (Class IIIa, 5 mW, 650 nm wavelength) with a maximum modulation frequency of 50 kHz. The lasers were controlled using a Teensy 4.0 development board capable of sub-microsecond digital output.

The measured rise and fall times of the laser diodes were $5.62 \pm 5.26 \mu\text{s}$ and $410.1 \pm 9.6 \mu\text{s}$, respectively. Overall, the measured latency between the computer-issued command and laser diode activation was sub-millisecond.

The viewing angles of the laser reflections on the objects were $0.16 \pm 0.08^\circ$ horizontally and $0.33 \pm 0.06^\circ$ vertically. The size of the reflections varied slightly depending on the optical properties of the objects. The viewing angle between the leftmost and right-most objects was 45.6° .

A computer screen was positioned 10 cm behind and above the objects, displaying eight pictograms corresponding to the objects. The pictograms were displayed on a BenQ XL2420Z monitor (24-inch, 16:9 aspect ratio, 1920 $\times$ 1080 resolution, 60 Hz refresh rate) using PsychoPy [6]. They were arranged in a 1 $\times$ 8 matrix. The viewing angle of a pictograms was 1.48° in both horizontal and vertical directions. The viewing angle from the leftmost to the right-most pictogram was 9.5° .

An optical sensor was attached to the top-left corner of the screen to capture precise stimulus onset timings for marker synchronization. Display presentation latency tests revealed an average delay of 3.1 ± 0.62 ms.

Stimuli: In this study, we used the same stimulus onset asynchrony (SOA) of 250 ms and inter-stimulus interval (ISI) of 150 ms as Kolkhorst et al. (2020) [5]. To reduce the complexity of the setup and the trial duration, we replaced the single sensor robotic arm with a parallel highlighting by multiple lasers.

As highlighting neighboring stimuli can lead to co-amplification of responses [7], two *non-adjacent* objects (laser condition) or pictograms (screen condition) were highlighted simultaneously during each highlight event, yielding a target-to-non-target ratio of 1:3. Although this target probability is slightly higher than the recommended range of 0.1–0.2 for eliciting P300 [8], it allowed us to reduce Kolkhorst et al.'s trial duration by 50%. While the amplitude of the P300 component is known to increase with longer target-to-target intervals (TTI) [8, 9] and a low target probability [10–13], excessively long TTIs and very low target probabilities may

also lead to participant fatigue and a loss of attention. To balance these factors, we used a variable TTI with a minimum interval of 500 ms. An example stimulation sequence is shown in Figure 2. The highlighting sequences were pseudo-randomly generated and satisfied three constraints: (1) two highlights of the same object or pictogram were at least 500 ms apart, (2) two highlighted objects or pictograms were not spatially adjacent, and (3) each object/pictogram was highlighted exactly 12 times. This produced highlighting sequences of a fixed length ($T = 48$ highlights, 12 s) with controlled target probability, variable and unpredictable TTIs, and reduced the co-amplification by neighboring stimuli.

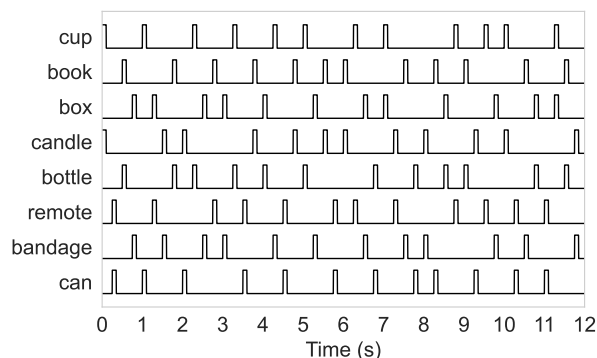


Figure 2: **Example stimulation sequence.** Each object and trial presented a unique pseudo-random sequence.

Experimental procedure: Before the experiment, the participants were familiarized with the task, object names, corresponding pictograms and the audio instructions which determined the target object or pictogram. Three practice trials were completed to make sure participants gazed at the correct object or pictograms.

Per trial, a participant's goal was to select an object placed on the table. There were 24 trials for each condition (6 for each of the objects) presented in a pseudo-random order. After 4 trials, the participants were allowed a self-paced break.

The target object's position was indicated through an audio cue (see Figure 1). It provided a logical explanation of the target object and emulated a real-world object search and selection. Additionally, the audio cue instructed to either gaze at the objects directly, or to the corresponding pictograms on the screen. The latter required participants to first find the object on the table, then focus their attention to the correct pictogram on the screen. The audio cue was repeated until participants pressed a button. It served to indicate that the target had been found, and initiated the stimulation after a random interval of 3–5 s. Participants were instructed to keep gazing at the target throughout the stimulation, which consisted of either a laser dot (see Figure 1b) or a highlighting of the pictogram (see Figure 1c).

Analysis: Data of three participants were excluded from data analysis. One participant withdrew from the experiment due to excessive eye discomfort. Data of two

participants were excluded because of technical problems related to the optical screen sensor in their session.

The remaining EEG data of 22 participants were band-pass filtered at 0.5–20 Hz using IIR forward-backward filtering. The data were then segmented to target and non-target epochs from 200 ms pre-stimulus to 600 ms post-stimulus. Epochs were baseline-corrected using the 250 ms pre-highlight interval. The epoched data were decimated to 250 Hz. Epochs with a peak-to-peak difference above $100 \mu\text{V}$ were rejected from analysis. In total, 15.9 % of epochs were excluded from further analysis.

We analyzed and compared the ERPs between the screen and laser conditions. Per participant, we estimated the amplitudes and latencies by region-of-interests (ROIs). The N1 ROI comprised of electrodes PO7, PO3, POz, PO4, PO8, PO9, O1, Oz, O2 and PO10, while the P2 ROI included FCz, Cz, CPz, Pz, POz and Oz. Amplitudes were defined as the minimum voltage within 0–250 ms for the N1 and the maximum voltage within 150–300 ms for the P2. Values were then averaged across electrodes within each ROI. To test for differences in ERP amplitudes and latencies, we used the two-sample pairwise t-test.

RESULTS

Target and non-target responses from both the screen and laser conditions were averaged across participants. The resulting event-related potentials (ERP) and corresponding scalp topographies are shown in Figure 3. Topographic maps were computed by averaging activity around the peak latency of each component. We observed clear differences in ERP waveforms and scalp activity patterns.

In the screen condition, the distinctive P1 component is observed between 112–132 ms with positive amplitudes over central electrodes and negative amplitudes over occipital electrodes. In the laser condition, a small P1 component is observed over central electrodes.

The N1 component is visible in both conditions but differences in the waveform morphology can be observed. In the screen condition, the N1 appears as a sharp negative peak, while in the laser condition it shows a broader negative deflection with a wider onset, which may suggest the presence of overlapping peaks. Additionally, the N1 amplitude appears larger in the laser condition compared to the screen condition.

The P2 component is visible in both conditions but its scalp activation patterns differ slightly. In the screen condition, P2 is primarily localized over parieto-occipital regions, with a slight extension toward central electrodes. In contrast, the laser condition shows P2 activity predominantly over central electrodes.

Temporal patterns also differ between the two conditions. In the screen condition, scalp activity initially appears in occipital regions, shifts toward central electrodes, and then returns toward occipital regions before gradually dissipating. In the laser condition, strong occipital activ-

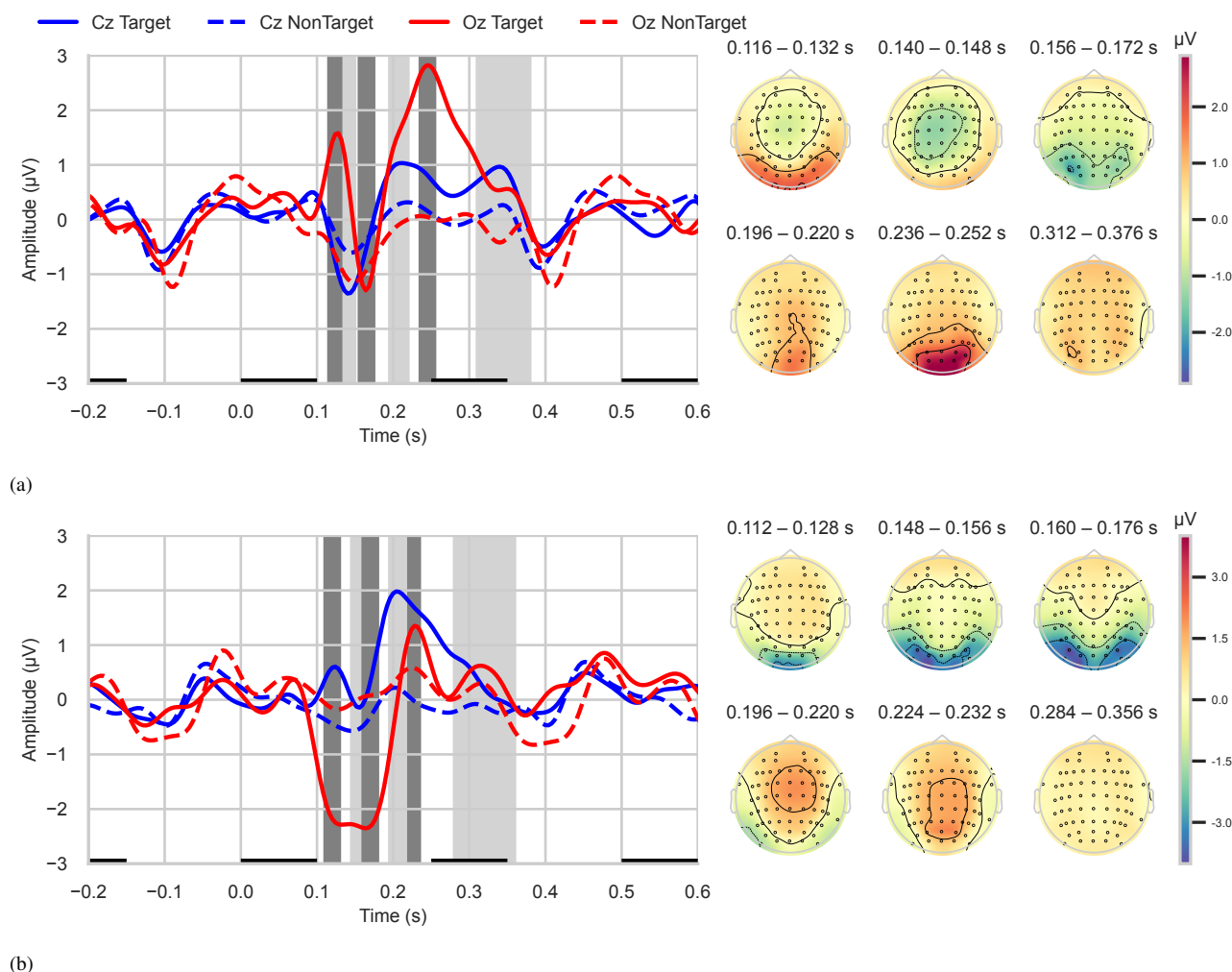


Figure 3: ERP timeseries and topoplots. Displayed are the grand average ERP timeseries (left) and topoplots (right) for the screen-based (a) and laser-based (b) conditions. Gray areas indicate the intervals across which the topoplots are shown. The black horizontal bars indicate highlighting intervals.

ity persists initially and subsequently shifts toward central electrodes before gradually dissipating.

For both conditions, no clearly distinguishable P3 component was observed despite the stimuli being designed to elicit a P300 component.

Component amplitude and latency analysis: Overall, ERP amplitudes observed in this study were lower than those typically reported for visual oddball paradigms. Figure 4 shows the N1 and P2 amplitudes and latencies across participants. Paired-samples t tests were conducted to compare ERP amplitudes and latencies between the laser and screen conditions. Effect sizes are reported as Cohen's d_z for paired designs. As shown in Table 1, the N1 amplitude was significantly more negative in the laser condition than in the screen condition ($M_{\text{diff}} = -1.47 \mu\text{V}$, 95% CI $[-2.35, -0.58]$, $t(21) = -3.44$, $p = .002$, $d_z = -0.73$). No significant condition differences were observed for N1 latency, P2 amplitude or P2 latency ($p > .05$). As mentioned in the methods section, we estimated the minimum and maximum values for ROIs instead of using only the Cz and Oz electrodes.

Our results suggest that the P2 responses are comparable between the screen and laser conditions despite differences in their scalp activation patterns.

DISCUSSION

This study compared ERP responses elicited by a conventional screen-based BCI and a scene-based BCI in which objects were highlighted using laser pointers.

The results indicate that the observed ERP waveforms differed between the two conditions. Although both conditions timing-wise used an identical stimulation protocol, the screen condition exhibited clearly distinguishable P1, N1, and P2 components, whereas the laser condition exhibited a broader N1 component followed by a reduced P2 component. One possible explanation for the missing P1 in the laser condition is that the broader N1 response may overlap with the P1 component, making it less distinguishable. In addition, differences were observed in the spatial distribution of the P2 component. In the screen condition, the P2 response was primarily localized over

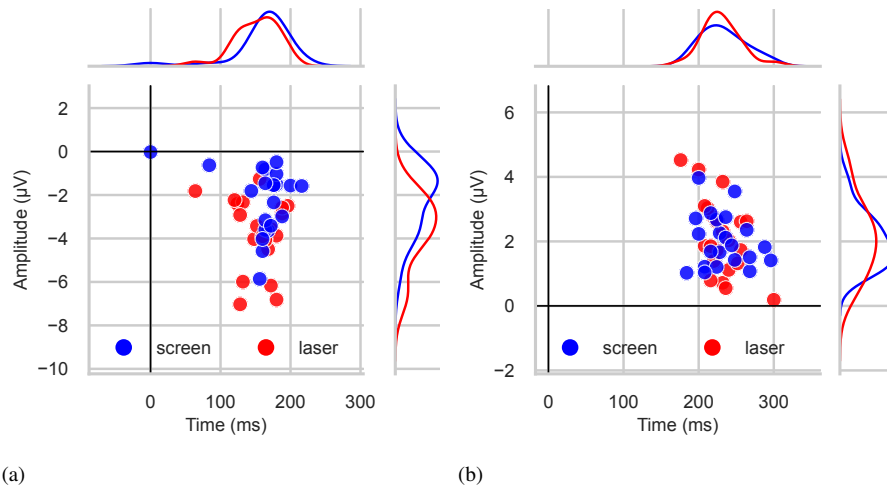


Figure 4: **Amplitude and latency distributions of ERP components.** Displayed are the ERP amplitude and latencies under laser and screen conditions. (a) N1 component amplitude and latency distribution. (b) P2 component amplitude and latency distribution.

Table 1: **Paired-samples t -tests comparing ERP measures between screen and laser conditions.** Note, M_{screen} and M_{laser} denote means for the screen and laser conditions, respectively. M_{diff} represents the mean paired difference (laser – screen). Confidence intervals correspond to the mean difference. Effect sizes are reported as Cohen’s d_z for paired designs.

Outcome	M_{screen} (SD)	M_{laser} (SD)	M_{diff}	95% CI	$t(21)$	p	d_z
N1 amplitude (μV)	-2.20 (1.51)	-3.66 (1.61)	-1.47	[-2.35, -0.58]	-3.44	.002	-0.73
N1 latency (ms)	160.55 (43.32)	151.82 (30.72)	-8.73	[-31.73, 14.28]	-0.79	.439	-0.17
P2 amplitude (μV)	2.02 (0.82)	2.13 (1.17)	0.11	[-0.33, 0.55]	0.53	.604	0.11
P2 latency (ms)	233.09 (30.06)	230.18 (25.74)	-2.91	[-19.38, 13.57]	-0.37	.717	-0.08

occipital regions, whereas in the laser condition it was most prominent over central electrodes. This shift in activation patterns may reflect differences in visual processing between controlled screen stimuli and stimuli presented in a more complex real-world environment.

We also observed a significant difference in N1 amplitude, which was larger in the laser condition than in the screen condition. This is consistent with findings reported by Kolkhorst et al. (2019) [4]. Assuming that the N1 component reflects early visual processing, this effect may be caused by differences in the visual characteristics of the stimuli. Despite being smaller, the laser dot may have appeared more intense or visually salient than the screen stimulus.

Although the stimulation protocol was based on an odd-ball paradigm, a clearly distinguishable P300 component was not consistently observed. One possible explanation is the relatively small visual angle of the stimuli, which may have reduced the signal-to-noise ratio of the P300 component. The waveforms provided in Figure 3 seem similar to those reported by the ISCEV standard for clinical visual evoked potentials [14, 15]. Specifically, at electrode Oz, the waveform of the laser condition resembles the pattern onset/offset VEP and the waveform of the screen condition resembles a flash VEP. However, as our protocols are substantially different from those used in the ISCEV standards, this similarity may not indicate a functional relationship and requires further investigation. Our data suggest that ERP responses in scene-based BCIs may differ in morphology and spatial distribution

from those observed in conventional screen-based systems. Consequently, decoding approaches for screen-based BCIs may need to account for these differences when applied to scene-based paradigms. This will be explored in future work.

Although the present study demonstrates that laser-based stimulation can elicit ERP responses with comparable intensities, it still needs to be tested, if in-the-scene BCIs can improve the usability or the workload in practical applications as this depends on factors beyond electrophysiological responses, including system complexity, setup time, the need for caregivers, user comfort, and the level of cognitive or physical workload (e.g., eye strain). Future work should investigate these aspects through dedicated usability studies.

In addition, the stimulation paradigm itself may influence usability and performance. The present study focuses on an ERP paradigm, but other BCI paradigms such as steady-state visually evoked potential (SSVEP) [16] or code-modulated visual evoked potential (c-VEP) [17] approaches may offer higher information transfer rates [18]. However, VEP-based paradigms are sensitive to stimulus properties such as size, color, luminance, and flicker frequency. In complex real-world environments, these properties may vary substantially, and their effects on VEP responses should be carefully investigated before adopting such paradigms in scene-based BCIs.

Finally, the hardware setup for scene-based BCIs may also play an important role. In the present work, objects were highlighted using laser pointers for its simplicity.

For practical applications, more advanced highlighting methods may be adopted.

CONCLUSION

This work presents an in-the-scene BCI using a laser-based implementation for highlighting real-world objects in parallel. EEG responses elicited by this approach were compared with those obtained using a conventional screen-based BCI. Statistical analysis of evoked ERP components revealed that the two conditions produced a comparable P2 amplitude and latency. While the N1 component did not show a significant difference in latency, a significantly larger N1 amplitude was found in the laser condition. These results suggest that laser-based highlighting elicits neural responses that are slightly different to those used in traditional ERP BCIs. Future work should investigate adapting decoding methods developed for screen-based paradigms for in-the-scene BCI systems.

ACKNOWLEDGEMENT

This project has received funding from the European Union's research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101118964.

REFERENCES

- [1] Farwell L, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*. 1988;70(6):510–523.
- [2] Wolpaw JR et al. Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. *Neurology*. 2018;91(3):e258–e267.
- [3] Pinegger A, Hiebel H, Wriessnegger SC, Müller-Putz GR. Composing only by thought: Novel application of the P300 brain-computer interface. *PLOS ONE*. 2017;12(9):e0181584.
- [4] Kolkhorst H, Veit J, Burgard W, Tangermann M. Heterogeneity of event-related potentials in a screen-free brain-computer interface. In: *Proceedings of the 8th Graz Brain-Computer Interface Conference 2019*. Graz, Austria, 2019, 228–233.
- [5] Kolkhorst H, Veit J, Burgard W, Tangermann M. A robust screen-free brain-computer interface for robotic object selection. *Frontiers in Robotics and AI*. 2020;7:38.
- [6] Peirce J et al. Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*. 2019;51(1):195–203.
- [7] Townsend G et al. A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clinical Neurophysiology*. 2010;121(7):1109–1120.
- [8] Duncan CC et al. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*. 2009.
- [9] Gonsalvez, Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*. 2002.
- [10] Farwell LA, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*. 1988;70(6):510–523.
- [11] Donchin E, Coles MGH. Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*. 1988;11(3):357–374.
- [12] Picton TW. The P300 Wave of the Human Event-Related Potential. *Journal of Clinical Neurophysiology*. 1992;9(4):456.
- [13] Polich J. P300 Clinical Utility and Control of Variability. *Journal of Clinical Neurophysiology*. 1998;15(1):14.
- [14] Odom JV et al. ISCEV standard for clinical visual evoked potentials: (2016 update). *Documenta Ophthalmologica*. 2016;133(1):1–9.
- [15] Habjan M Šuštar et al. ISCEV standard for clinical visual evoked potentials (2025 update). *Documenta Ophthalmologica*. 2025;151(2):97–112.
- [16] Vialatte FB, Maurice M, Dauwels J, Cichocki A. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in Neurobiology*. 2010;90(4):418–438.
- [17] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. *Journal of Neural Engineering*. 2021.
- [18] Neeling M de, Van Hulle MM. Single-paradigm and hybrid brain computing interfaces and their use by disabled patients. *Journal of Neural Engineering*. 2019;16(6):061001.